

Artificiële intelligentie tot ondersteuning van de assetconfiguratie

26 februari 2021, 01:00

Om een zonnepanelenpark met een online monitoring platform te verbinden, is een accurate configuratie nodig. Deze tijdrovende, handmatige en foutgevoelige ingreep kan worden ondersteund door een datagedreven aanpak.

Met het oog op een maximale duurzaamheid helpen de experts en ingenieurs van het Brusselse bedrijf 3E verschillende industriële actoren bij de optimalisatie van de productie en het gebruik van (hernieuwbare) energieën, via innovatieve diensten van softwareontwikkeling en consultancy. 3E kan bogen op een lange geschiedenis van innovatie op het vlak van hernieuwbare-energie-installaties, slimme stroomvoorzieningen en energiebewuste procesverbeteringen overal ter wereld.

Het SynaptiQ-platform van 3E is bedoeld voor de monitoring van hernieuwbare energiebronnen, zoals windturbine- en zonnepanelenparken. Een van de uitdagingen bij de verbinding van een zonnepanelenpark met SynaptiQ is de zogenaamde onboarding, of de configuratie van het data-verzamelingsproces voor die installatie en de correcte configuratie van de digitale kopie ervan om tot een exacte berekening van de kritieke prestatie-indicatoren (KPI's) te komen. Afhankelijk van de grootte, de complexiteit en de specifieke eigenschappen van het park kan deze onboarding een tijdrovend proces zijn waarvoor ervaren ingenieurs moeten worden ingezet. De belangrijkste moeilijkheden voor het engineeringteam zijn:

- het gebrek aan standaardisatie met een ruime waaier van hardwarefabrikanten, communicatieprotocollen, dataformaten en ad-hoc configuraties van SCADA-systemen ter plaatse
- vaak onvolledige (of totaal ontbrekende) documentatie over de topologie, de types van apparatuur en de interconnecties binnen een park

Voor het configureren van de grotere parken zijn dus vaak een uitvoerig onderzoek en een zorgvuldige analyse van de verzamelde data nodig, die gespreid is over honderden apparaten, met het oog op de reverse-engineering van de configuratie ervan. Door de grote vraag naar nieuwe te verbinden centrales betekent dit dat soms alleen de belangrijkste prestatie-indicatoren worden geconfigureerd, terwijl indicatoren die minder relevant zijn worden verwaarloosd, vooral als de configuratiekosten hoog zijn.

De Synaptiq-software heeft reeds een zeer grondig foutopsporingssysteem aan boord dat permanent ongekende en onjuiste indicatoren van een park registreert (bijv. discrepanties tussen de ontvangen data en de configuratie van de digitale kopie), wat bijzonder nuttig is voor het engineeringteam.

Om nog meer voordeel te halen uit deze foutinformatie, werd een onderzoeksproject opgezet in samenwerking met het EluciDATA Lab van Sirris. Het project strekt tot doel de configuratietijd in te korten en bijgevolg het bereik van de indicatoren te vergroten, via een datagedreven aanpak die voor een specifiek zonnepanelenpark de configuratieparameters en hun overeenkomstige waarden aanbeveelt.

Zorgvuldige voorbereiding van de data is een must

De door 3E verstrekte gegevens bevatten een willekeurige populatie van enkele duizenden zonnepanelenparken, waarvan sommige sterk op elkaar lijken en andere zeer van elkaar verschillen. De variabiliteit van elke configuratie gaat van een paar tot tientallen parameters: voor bepaalde parken volstaan standaard parameters en waarden, terwijl voor andere meer dan 10 parameters worden gespecificeerd.

Belangrijk om weten is dat configuraties in zeer verschillende mate voorkomen. Sommige komen slechts één maal voor, terwijl andere de dataset domineren met meer dan 2500 voorkomens. Het is dan ook belangrijk dat we eerst de verhouding van de eenmalige configuraties afwegen zodat we nog steeds een groter gewicht kunnen toekennen aan frequent voorkomende configuraties, zonder echter die met een laag aantal voorkomens te verwaarlozen. Daartoe berekenen we het aantal m per configuratie:

$$m = \begin{cases} 2, & \text{if } \log_2(\#n) < 2 \\ 100, & \text{if } \log_2(\#n) > 100 \\ \log_2(\#n), & \text{else} \end{cases}$$

dat het gewicht per configuratie bepaalt en waarbij n het aantal voorkomens van een configuratie is.

Associatieregels afleiden

Om een correcte configuratie aan te bevelen, hanteren we zogenaamde 'association rules'. Een eerste stap naar de voorspelling van een configuratie bestaat erin zogenaamde 'frequent itemsets' te berekenen op basis van de afgewogen set van configuraties, die de verschillende parameters en waarden groeperen in lijsten van items die (meer) frequent samen voorkomen. Het aantal maal dat de eenmalige parameterinstellingen samen voorkwamen in verhouding tot de volledige dataset, wordt weergegeven door de support. Hoe hoger de *support*, hoe duidelijker het samen voorkomen van meerdere parameters en hun overeenkomstige waarden. Het resultaat is een lijst van meer dan 18.000 frequente items en hun support.

	support	itemsets
1367	0.240602	(csv.timestamp.column__1, csv.separator__, serialid.pattern.group__1)
4611	0.159398	(csv.header.line__1, csv.timestamp.column__1, csv.separator__)
180	0.154887	(timestamp.pattern__yyyy-MM-dd HH:mm:ss, csv.timestamp.column__1, csv.separator__)
178	0.132331	(timestamp.pattern__yyyy-MM-dd HH:mm:ss, csv.timestamp.column__1, serialid.pattern.group__1)
4612	0.120301	(csv.header.line__1, csv.timestamp.column__1, serialid.pattern.group__1)
...	...	...

In een tweede stap passen we het algoritme voor de extractie van associatieregels toe op de eerder verkregen set van frequente parameterinstellingen, met de bedoeling regels vast te stellen op basis van gangbare configuraties. Het resultaat is een set van regels die aangeven welke configuratie-instellingen bij elkaar horen. Voor elke regel kunnen meerdere *geschiktheidsmetingen* worden uitgevoerd, met andere woorden metingen van de sterkte van een verhouding. Behouden we alleen de relevante regels, dan blijven er nog steeds meer dan 700.000 over.

Vertrekkend van deze regels kunnen dan configuratievoorspellingen worden afgeleid:

Input settings	recommendations	goodness measures				
antecedents	consequents	antecedent support	consequent support	support	confidence	lift
(csv.header.line__1, csv.timestamp.column__1, dynamic.units__false, tag.pattern.group__3)	(csv.data.line__2, serialid.pattern.group__2)	0.006015	0.006015	0.006015	1.0	106.25
(csv.header.line__1, csv.timestamp.column__1, csv.data.line__2, serialid.pattern.group__2)	(dynamic.units__false, tag.pattern.group__3)	0.006015	0.006015	0.006015	1.0	106.25
(csv.header.line__1, dynamic.units__false, serialid.pattern.group__2, csv.data.line__2)	(tag.pattern.group__3, csv.separator__)	0.006015	0.006015	0.006015	1.0	106.25
(csv.header.line__1, dynamic.units__false, tag.pattern.group__3, csv.separator__)	(csv.data.line__2, serialid.pattern.group__2)	0.006015	0.006015	0.006015	1.0	106.25
(csv.header.line__1, csv.data.line__2, serialid.pattern.group__2, csv.separator__)	(dynamic.units__false, tag.pattern.group__3)	0.006015	0.006015	0.006015	1.0	106.25

Resultatenbeoordeling

Het doel was aanbevelingen te verstrekken voor de parken waarvoor fouten bij de data-import werden waargenomen. Dergelijke fouten kunnen drie oorzaken hebben:

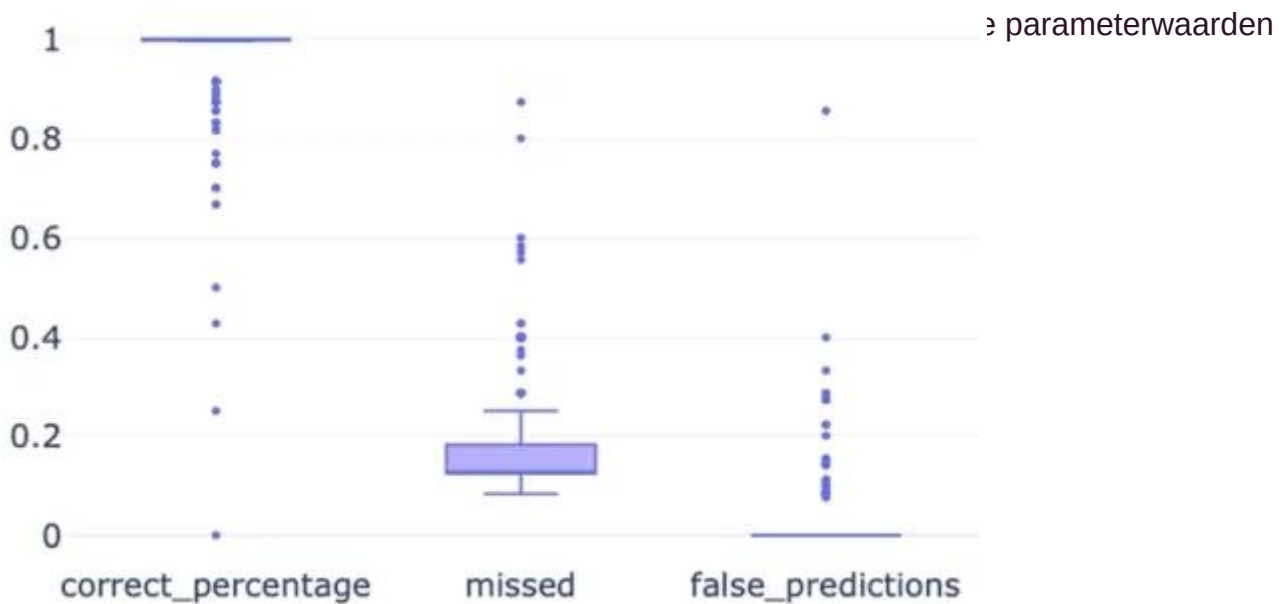
- Een of meer parameterwaarden zijn verkeerd ingesteld.
- Een of meer parameterinstellingen ontbreken.
- Een of meer parameterinstellingen zijn overbodig en misleidend.

Net daarom werden per park meerdere subsets van parameters en hun overeenkomstige waarden gebruikt als inputvariabelen voor de aanbevelingsmotor. De subsets bestaan uit alle combinaties van 2 tot $k - 1$ parameter-waardeparen, waarbij k het aantal parameters is dat in de aanvankelijke configuratie was bepaald. Het aantal subsets neemt exponentieel toe met k en, afhankelijk van de inputparameters, kunnen verschillende parameters en voor elke parameter verschillende waarden per subset worden aanbevolen. Om de definitieve parameters en parameterwaarden vast te leggen, werd het principe van de meerderheidsstemming toegepast: de verzameling van parameters en de overeenkomstige waarde die het meest werd aanbevolen in alle subsets, werd geselecteerd. Bij gelijke stand werd gekozen voor de set met de hoogste betrouwbaarheid.

Over het algemeen werden goede resultaten verkregen, met:

- een gemiddeld percentage van correct voorspelde parameterwaarden van ongeveer 90%

- een gemiddeld percentage van minder dan 20% van gemiste parameters, d.w.z. parameters die niet voorkomen



Conclusie

Uit de resultaten blijkt dat een aanpak op basis van artificiële intelligentie effectief kan worden gehanteerd tot ondersteuning van de assetconfiguratie: we kunnen aanbevelingen geven over welke configuratieparameters moeten worden verstrekt en wat de waarde ervan moet zijn.

Dit zijn evenwel de belangrijkste aandachtspunten:

- Een zorgvuldige voorverwerking van de data is een must om tot goede resultaten te komen: de dataset als dusdanig in het algoritme invoeren zou resulteren in (foute) aanbevelingen die worden gedomineerd door meer frequent voorkomende configuraties.
- Er moet absoluut een specifieke procedure worden vastgelegd voor de selectie van de aanbeveling met de hoogste betrouwbaarheid: het algoritme voor de extractie van regels genereert een overvloed aan regels, wat tot tegenstrijdige aanbevelingen kan leiden.